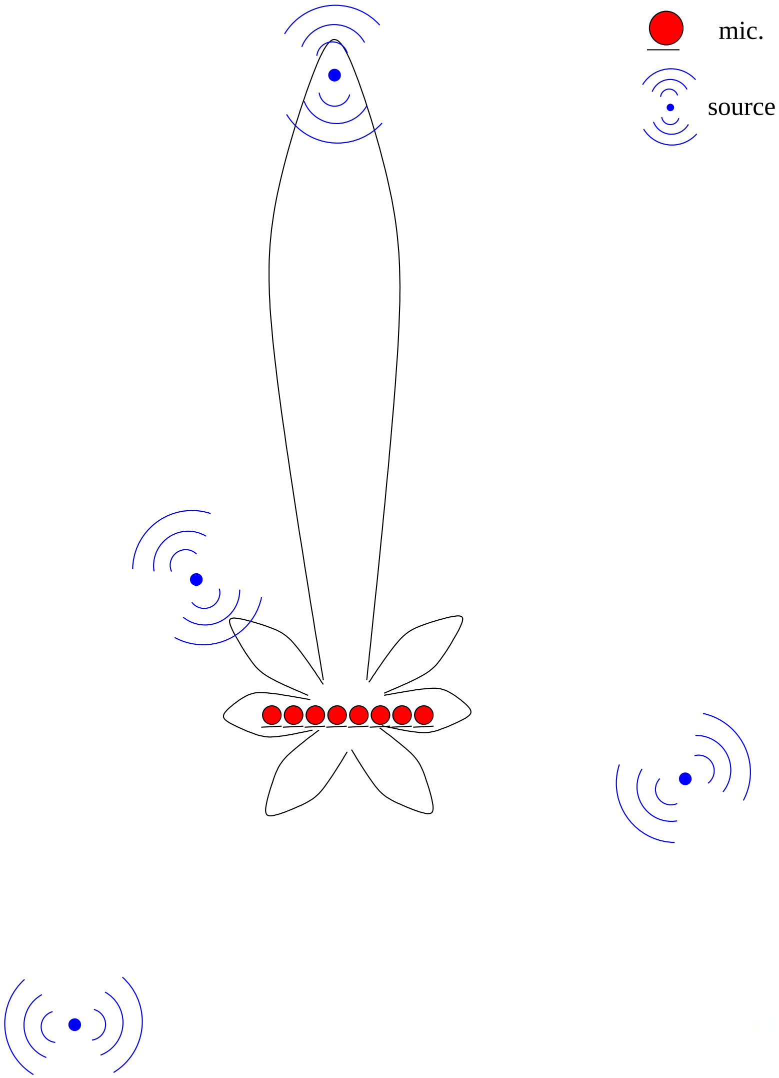
Recovering Audio Sources in a multi-path Environment

Lucas Parra, Paul Sajda

Air-coupled Acoustic Sensors Workshop

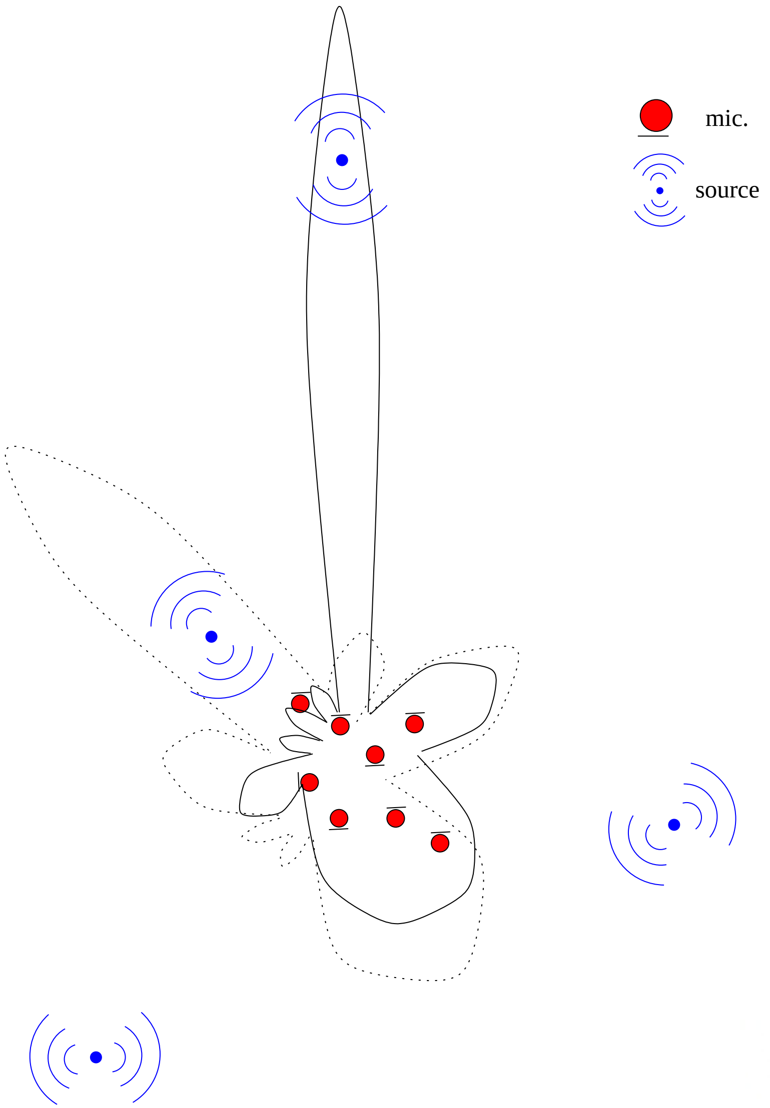
Geometric (Adaptive) Beamforming

- What it is
  - microphone array with fixed geometric configuration
  - adaptive algorithms to steer and adjust beam pattern
- Typical Applications
  - Attenuation of jammers
  - source localization
- Problems/Issues
  - Requires know/fixed array configuration
  - Cannot handle multiple sources
  - Signal leakage, reverberation



Statistical (Blind) Beamforming

- What it is
  - Multiple sensors at arbitrary locations
  - adaptive algorithm to recover independent/decorrelated source signals
- Typical Applications
  - simultaneous recovery of multiple sources
  - jammer attenuation under reverberation, and target signal leakage
- Problems/Issues
  - requires low noise sensors
  - computational complexity



Recovering Speech from Simultaneous Recording (Statistical Beamforming Demo)

... isolating individual speakers with multiple microphones ...

- Instantaneous mixture - corresponds to environment with no reverberation and known time delays.
- Solution can't assume knowledge of speaker/source location - requires a "blind" algorithm.
- Demo 1 - linear mixing of 10 speakers.
- Demo 2 - "real-world" deconvolution with 2 speakers.
- Sarnoff algorithm exploits non-stationarity of speech signal, performing multiple decorrelation across time to compute a matrix of "unmixing" FIR filters.

The Problem: Convolutive Mixture

Acoustic signals x(t) recorded simultaneously in a reverberant environment A(τ) can be described as sums of differently convolved sources s(t).

$$x(t) = \sum_{\tau=0}^{P} A(\tau) s(t-\tau) + n(t)$$

with *dim*(x)≥ *dim*(s)

SARNOFF Corporation

Context on Blind Source Separation

PCA:

$$x = R s$$

$$R : s = R^T x \qquad <s_i s_j> = \delta_{ij} \lambda_i$$

ICA:

$$x = A s$$

$$A : s = A^{-1} x \qquad <s^n_i \, s^m_j> = \delta_{ij} \lambda_i^{n+m}$$

$$W : s = W x \qquad <s_i(0) \, s_j(t)> = \delta_{ij} \lambda_i(t)$$

BSS:

$$x = A \otimes s(t)$$

$$A : s(t) = A^{-1} \otimes x(t) \qquad < s^n_i(0) s^m_j(t)> = \delta_{ij} \lambda_{inm}(t)$$

$$W : s(t) = W \otimes x(t) \qquad < s_i(t) \, s_j(t')> = \delta_{ij} \lambda_i(t,t')$$

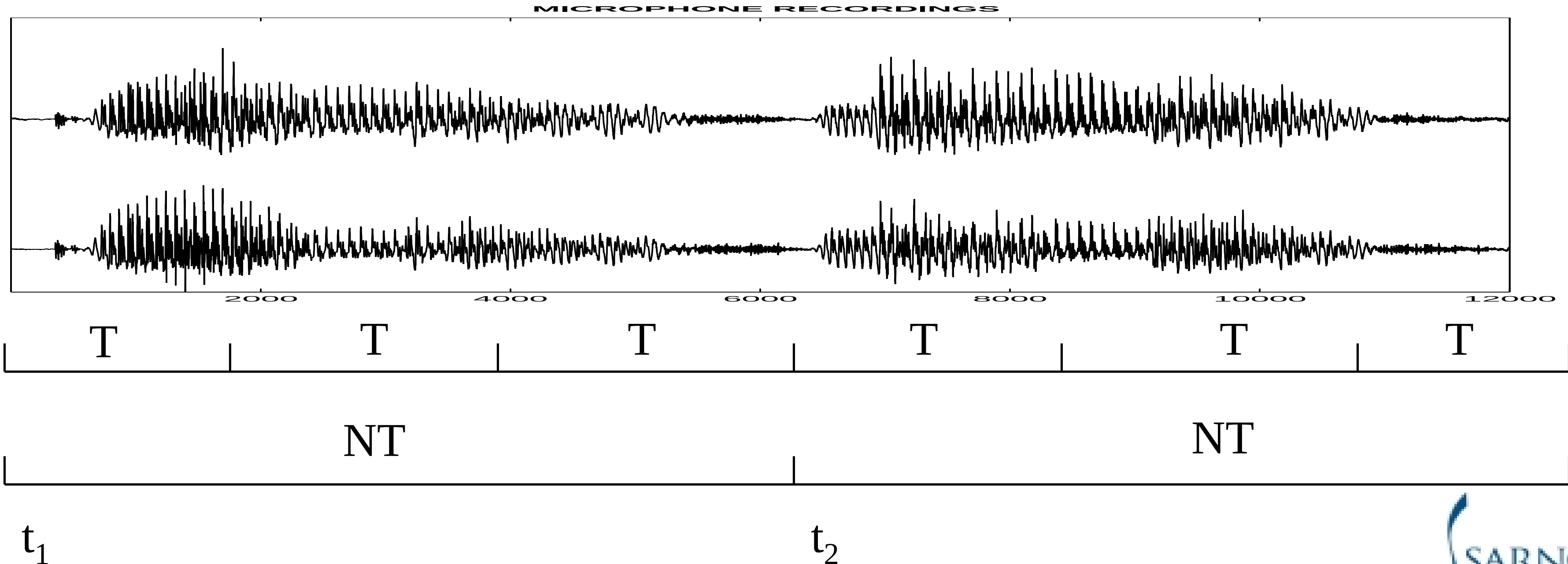
SARNOFF Corporation

Approach - Use Non-stationarity

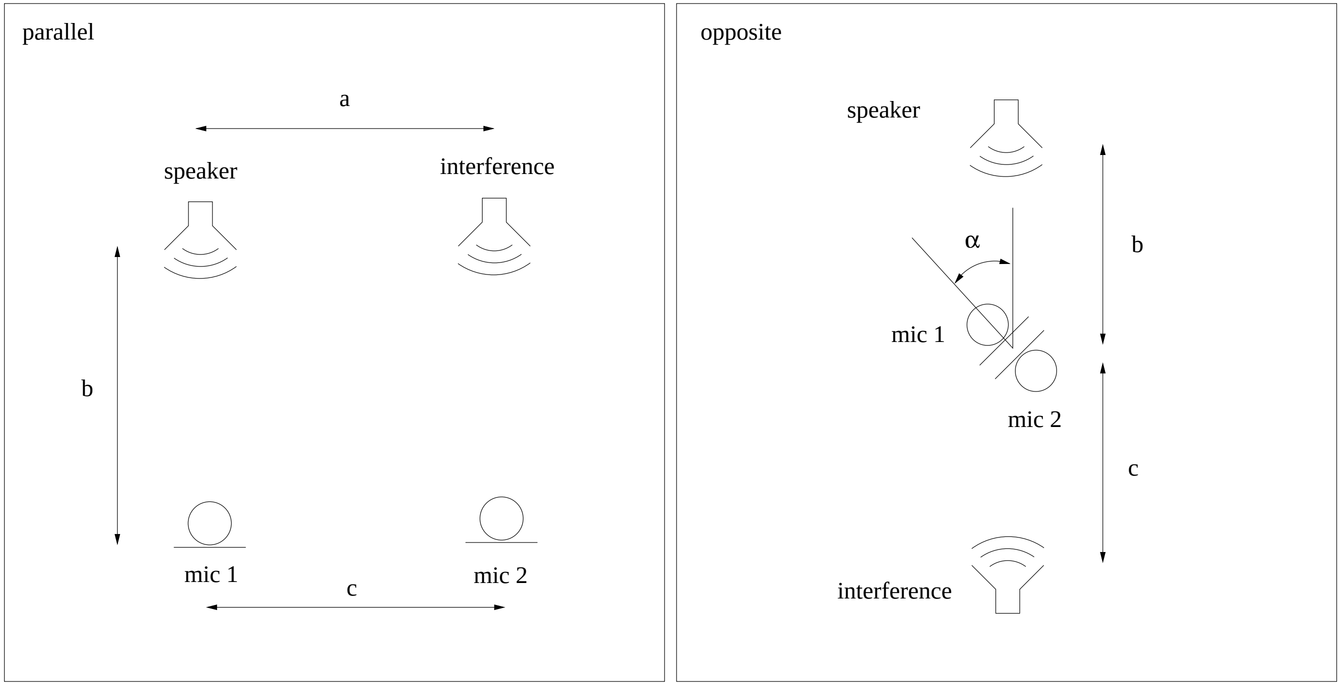
Measure time dependent second order statistic

$$\bar{R}_{\div}(\omega,t) = \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{x}(\omega,t+nT)\,\mathbf{x}^{H}(\omega,t+nT)$$

Where $\mathbf{x}(\omega,t)$ are the frequency components of frame $[\mathbf{x}(t), \ldots, \mathbf{x}(t+T)]$



Experimental Setup: Speaker with Interfering Source



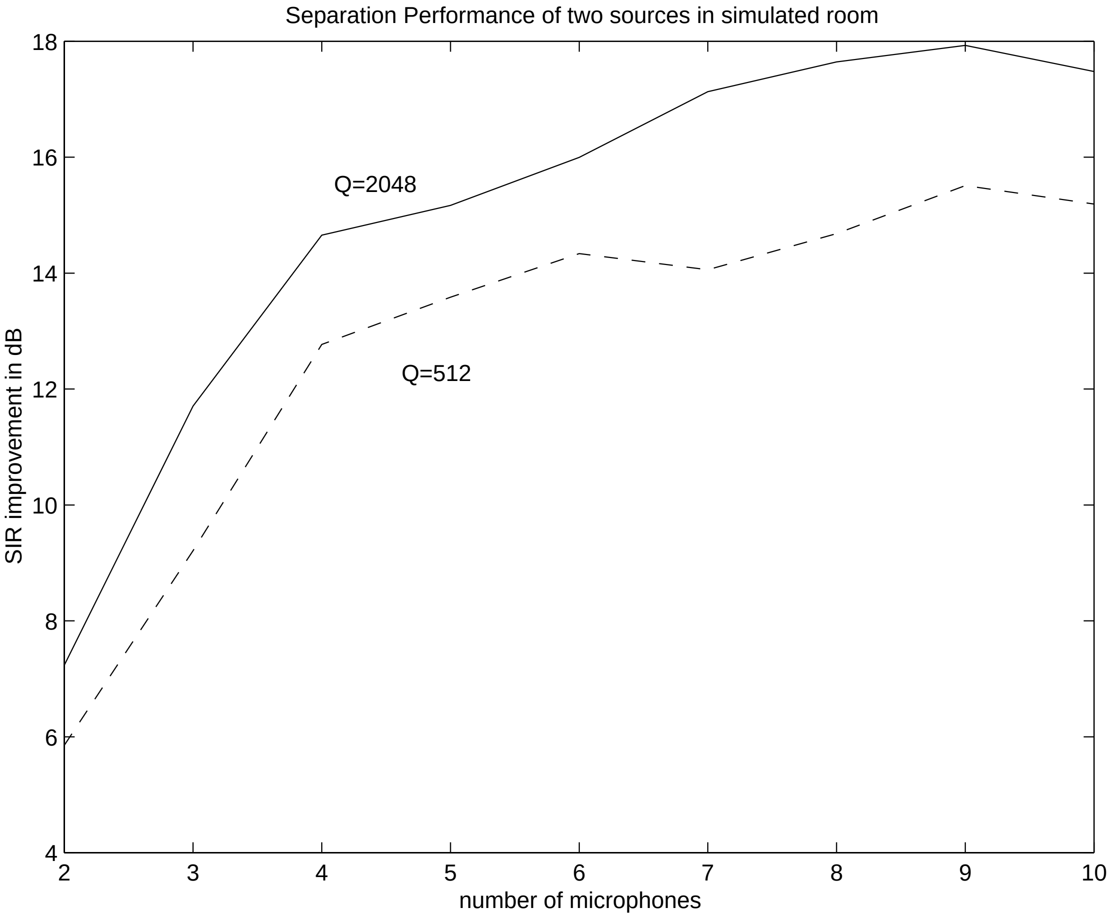
Reverberant environment (small office room). The interfering signal was a competing speaker or music.

Left: a = b =50? , c = 50? , 6? .

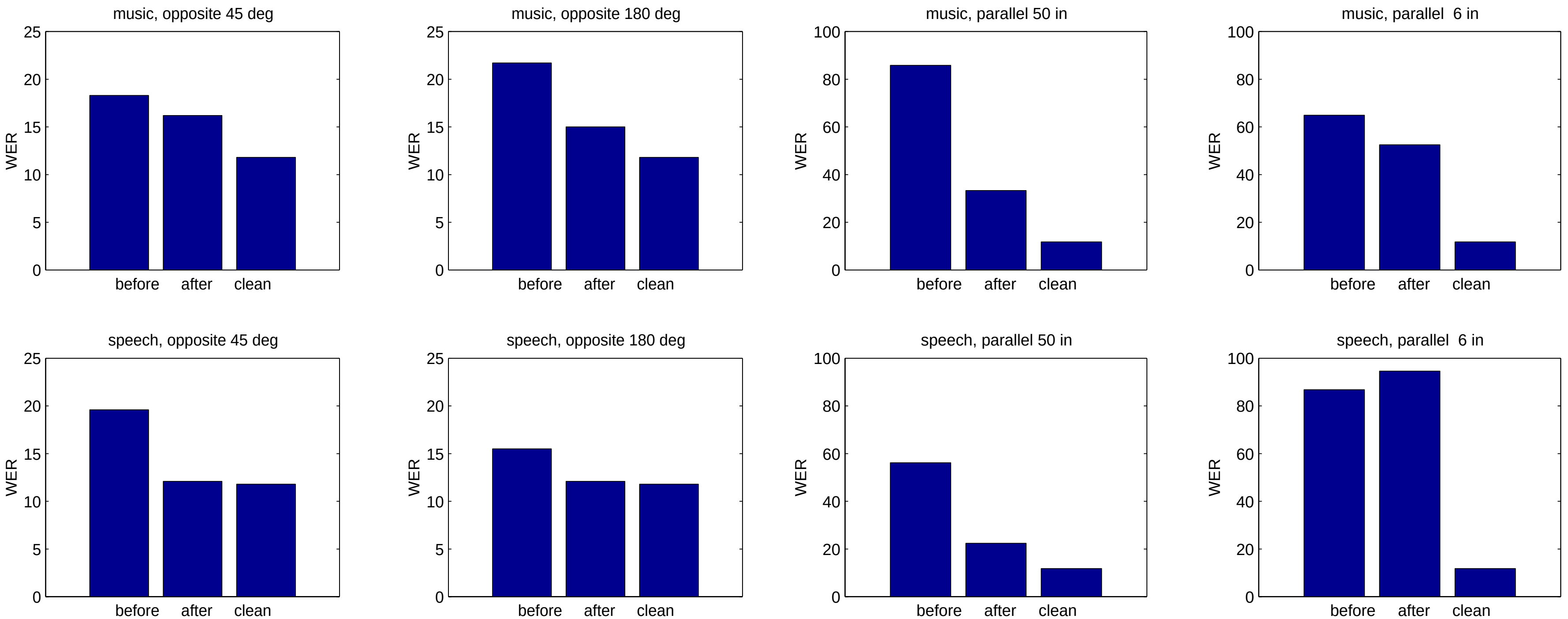
Right: b = 30? , c = 60? , ∝ = 45° , 180° .

Multiple Microphone Performance

Performance of multiple microphones in simulated room (small office) for separating a speaker from music background. Microphone distance 2m.



Speech Recognition Improvement



Word error rate (WER) of ViaVoice (IBM) on a short text (Wallstreet Journal article of 760 words length) before and after source separation. The result is contrasted to clean recording with no interfering source.

up to **50% reduction in word error rate** for IBM Viavoice